

# A note on prediction bias for state space models with estimated parameters

Magda Monteiro<sup>\*,†</sup> and Marco Costa<sup>\*,\*\*</sup>

<sup>\*</sup>ESTGA - Escola Superior de Tecnologia e Gestão de Águeda, Universidade de Aveiro, Portugal

<sup>†</sup>CIDMA - Centro de Investigação e Desenvolvimento em Matemática e Aplicações da Universidade de Aveiro

<sup>\*\*</sup>CMAF - Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa

## Abstract.

This paper aims to discuss some problems on state space models with estimated parameters. While existing research focus on the prediction mean squared error, this work presents some results on bias propagation into forecast and filter predictions when the mean vector of the state is taking with an estimation bias, namely, non recursive analytical expression for them. In particular, it is discussed the impact of mean bias in invariant state space models.

**Keywords:** state space model, prediction bias, Kalman filter, stationary state

**PACS:** 02.50.Ey

## INTRODUCTION

Once the model is placed in the state space form, the most usual algorithm to predict the latent vector is the Kalman filter (KF) algorithm. This algorithm is a procedure for computing, at each time  $t$  ( $t = 1, 2, \dots$ ), the optimal estimator of the state vector based on the available information until  $t$  and its success lies on the fact that is an online estimation procedure. Indeed, the main goal of the KF algorithm is to find estimates for the unobservable variables based on observable variables related to each other through a set of equations forming the state space model. According to the objectives of this work, and to simplify the explanation will be considered a class of state space models with a stationary state vector. Such models are defined by the equations:

$$\mathbf{Y}_t = \mathbf{H}_t \boldsymbol{\beta}_t + \mathbf{e}_t \quad (1)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t. \quad (2)$$

The general state space form applies to a multivariate time series  $\mathbf{Y}_t$ , a  $n \times 1$  vector of observable variables, which are related, via *measurement equation* (1), with the  $m \times 1$  vector of unobservable variables,  $\boldsymbol{\beta}_t$ , known as the state vector.  $\mathbf{H}_t$  is the system matrix,  $n \times m$ , of known coefficients and  $\mathbf{e}_t$  is a white noise  $n \times 1$  vector, called measurement error, of serially uncorrelated disturbances with mean zero and covariance matrix  $\boldsymbol{\Sigma}_e = E(\mathbf{e}_t \mathbf{e}_t')$ . Although the elements of  $\boldsymbol{\beta}_t$  are not observable, they are generated by a first-order Markov process according to (2), the *transition equation*. It is assumed that the state vector  $\boldsymbol{\beta}_t$  is a stationary VAR(1) process with mean  $E(\boldsymbol{\beta}_t) = \boldsymbol{\mu}$  and transition matrix  $\boldsymbol{\Phi}$  with all eigenvalues inside the unit circle, i.e.,

$$|\lambda_i(\boldsymbol{\Phi})| < 1 \text{ for all } \lambda_i \text{ such that } |\boldsymbol{\Phi} - \lambda_i \mathbf{I}| = \mathbf{0}. \quad (3)$$

The error  $\boldsymbol{\varepsilon}_t$  is a white noise vector,  $m \times 1$ , with mean zero and covariance matrix  $\boldsymbol{\Sigma}_\varepsilon = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ . The disturbances  $\mathbf{e}_t$  and  $\boldsymbol{\varepsilon}_t$  are assumed to be uncorrelated, that is,  $E(\mathbf{e}_t \boldsymbol{\varepsilon}_s') = \mathbf{0}$  for all  $t$  and  $s$ . The specification of the state space model is completed by the assumption that the initial state vector  $\boldsymbol{\beta}_0$  has a mean of  $\boldsymbol{\mu}$  and a covariance matrix  $\boldsymbol{\Sigma}$ , that is  $E(\boldsymbol{\beta}_0) = \boldsymbol{\mu}$  and  $\text{var}(\boldsymbol{\beta}_0) = \boldsymbol{\Sigma}$ , where  $\boldsymbol{\Sigma}$  is the solution of the equation  $\boldsymbol{\Sigma} = \boldsymbol{\Phi} \boldsymbol{\Sigma} \boldsymbol{\Phi}' + \boldsymbol{\Sigma}_\varepsilon$ .

## THE KALMAN FILTER ALGORITHM

Briefly, the Kalman filter is an iterative algorithm that produces, at each time  $t$ , an estimator of the state vector  $\boldsymbol{\beta}_t$  which is given by the orthogonal projection of the state vector onto the observed variables up to that time.

Considering the model (1)-(2), let  $\boldsymbol{\beta}_{t|t-1}$  denote the estimator of  $\boldsymbol{\beta}_t$  based on the observations  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}$  and let  $\mathbf{P}_{t|t-1}$  be its covariance matrix, i.e., the MSE matrix. Since the orthogonal projection is a linear estimator, the predictor for the observable vector  $\mathbf{Y}_t$  is given by  $\mathbf{Y}_{t|t-1} = \mathbf{H}_t \boldsymbol{\beta}_{t|t-1}$ .

When, at time  $t$ ,  $\mathbf{Y}_t$  is available, the prediction error or *innovation*,  $\boldsymbol{\eta}_t = \mathbf{Y}_t - \mathbf{Y}_{t|t-1}$ , is used to update the estimate of  $\boldsymbol{\beta}_t$  (*filtering*) through the equation

$$\boldsymbol{\beta}_{t|t} = \boldsymbol{\beta}_{t|t-1} + \mathbf{K}_t \boldsymbol{\eta}_t, \quad (4)$$

where  $\mathbf{K}_t$  is called the Kalman gain matrix and is given by

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \boldsymbol{\Sigma}_e)^{-1}.$$

Furthermore, the MSE of the updated estimator  $\boldsymbol{\beta}_{t|t}$ , represented by  $\mathbf{P}_{t|t}$ , verifies the relationship  $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}$ . On the other hand, at time  $t$ , the forecast for the state vector  $\boldsymbol{\beta}_{t+1}$  is given by the equation

$$\boldsymbol{\beta}_{t+1|t} = \boldsymbol{\mu} + \Phi(\boldsymbol{\beta}_{t|t} - \boldsymbol{\mu}) \quad (5)$$

and its MSE matrix is  $\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t} \Phi' + \boldsymbol{\Sigma}_e$ . For more details on Kalman filter algorithm see [3, 7].

## STATE SPACE MODELS WITH ESTIMATED PARAMETERS

When the disturbances  $\mathbf{e}_t$  and  $\boldsymbol{\varepsilon}_t$  are normally distributed the KF estimators minimises the MSE when the expectation is taken over all the variables because, in the normality case,

$$\boldsymbol{\beta}_{t|t} = E(\boldsymbol{\beta}_t | \mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots). \quad (6)$$

Thus, the conditional mean estimator is the minimum mean square estimator of  $\boldsymbol{\beta}_t$  and it is unbiased in the sense that the expectation of the estimation error is zero ([3]). However, when the disturbances are not normally distributed the equation (6) does not hold but, as orthogonal projection, the KF estimators are the minimum mean square linear estimators. Nevertheless, when the real parameters of the state space model  $\boldsymbol{\Theta} = \{\boldsymbol{\mu}, \Phi, \boldsymbol{\Sigma}_e, \boldsymbol{\Sigma}_\varepsilon\}$  are, for instance, substituted by their maximum likelihood (or other) estimates,  $\hat{\boldsymbol{\Theta}}$ , the theoretical properties of KF estimators are no longer valid. The usual approach in the analysis of the effects (implications) of applying estimates rather than using true values is the computation of the mean squared errors  $\mathbf{P}_{t|t}$  or  $\mathbf{P}_{t|t-1}$  taking into account that substitution. This approach is discussed in the literature, for instance in [2, 4] or more recently in [6] and it relies on the fact that substituting the model parameters by their estimates in the theoretical MSE expression that assumes known parameters values results in underestimation of the true MSE.

Indeed, denoting by  $\boldsymbol{\beta}_{t|t}(\hat{\boldsymbol{\Theta}})$  the predictor obtained from  $\boldsymbol{\beta}_{t|t}$  defined in (4), and analogously to the forecast predictor, by substituting  $\boldsymbol{\Theta}$  by  $\hat{\boldsymbol{\Theta}}$ , the MSE of the prediction error is

$$\begin{aligned} \text{MSE}_{t|t} &= E \left\{ \left[ \boldsymbol{\beta}_{t|t}(\hat{\boldsymbol{\Theta}}) - \boldsymbol{\beta}_t \right] \left[ \boldsymbol{\beta}_{t|t}(\hat{\boldsymbol{\Theta}}) - \boldsymbol{\beta}_t \right]' \right\} \\ &= \mathbf{P}_{t|t} + E \left\{ \left[ \boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t}(\hat{\boldsymbol{\Theta}}) \right] \left[ \boldsymbol{\beta}_{t|t} - \boldsymbol{\beta}_{t|t}(\hat{\boldsymbol{\Theta}}) \right]' \right\} \end{aligned} \quad (7)$$

Usually, the existent literature investigates methodologies to the second parcel of (7), that is, the contribution to the  $\text{MSE}_{t|t}$  resulting from 'parameters uncertainty'. In [4] it is suggested the application of Monte Carlo techniques combining with maximum likelihood estimation. Bootstraps procedures are applied to both parametric and nonparametric methods in [6].

## FORECAST AND FILTER BIAS

This paper intends initiating a research on the bias induced by taking parameters estimates instead of the true values.

As starting point of this research, it will be considered a state space model (1)-(2) where it will be admitted that all parameters are known except the mean vector  $\boldsymbol{\mu}$  that is estimated with a bias, i.e.,

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\lambda}, \quad (8)$$

where  $\boldsymbol{\lambda}$  is the estimation error.

As Kalman filter estimators are linear, the estimation error of  $\boldsymbol{\mu}$  will influence them additively. Indeed, the estimation error of  $\boldsymbol{\mu}$  induces bias recursively as follow. The initial state vector  $\hat{\boldsymbol{\beta}}_{1|0}$  has a bias equal to  $\boldsymbol{\lambda}$ , i.e.,  $\hat{\boldsymbol{\beta}}_{1|0} = \boldsymbol{\beta}_{1|0} + \text{bias}(\hat{\boldsymbol{\beta}}_{1|0}) = \boldsymbol{\beta}_{1|0} + \boldsymbol{\lambda}$ . The bias induced in forecast of  $\mathbf{Y}_t$  is given by

$$\hat{\mathbf{Y}}_{t|t-1} = \mathbf{H}_t \hat{\boldsymbol{\beta}}_{t|t-1} = \mathbf{Y}_{t|t-1} + \mathbf{H}_t \text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}) \quad (9)$$

which induces a bias in the filtering stage, namely,

$$\hat{\boldsymbol{\beta}}_{t|t} = \hat{\boldsymbol{\beta}}_{t|t-1} + \mathbf{K}_t (\mathbf{Y}_t - \hat{\mathbf{Y}}_{t|t-1}) = \boldsymbol{\beta}_{t|t} + (\mathbf{I}_m - \mathbf{K}_t \mathbf{H}_t) \text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}). \quad (10)$$

The bias of the state forecast as the form

$$\hat{\boldsymbol{\beta}}_{t|t-1} = \hat{\boldsymbol{\mu}} + \Phi(\hat{\boldsymbol{\beta}}_{t|t} - \hat{\boldsymbol{\mu}}) = \boldsymbol{\beta}_{t|t-1} + (\mathbf{I}_m - \Phi)\boldsymbol{\lambda} + \Phi \text{bias}(\hat{\boldsymbol{\beta}}_{t-1|t-1}). \quad (11)$$

These results can be summarized through recursive equations in terms of bias

$$\text{bias}(\hat{\boldsymbol{\beta}}_{1|0}) = \boldsymbol{\lambda}, \text{bias}(\hat{\boldsymbol{\beta}}_{t|t}) = (\mathbf{I}_m - \mathbf{K}_t \mathbf{H}_t) \text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}), \text{ and } \text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}) = (\mathbf{I}_m - \Phi)\boldsymbol{\lambda} + \Phi \text{bias}(\hat{\boldsymbol{\beta}}_{t-1|t-1}). \quad (12)$$

These equations allow obtaining non recursive analytical expressions, proved by mathematical induction, for forecast and filter bias, namely,

$$\text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}) = \left[ (\mathbf{I}_m - \Phi) + \left( \sum_{k=1}^{t-2} \prod_{i=1}^k \Phi (\mathbf{I}_m - \mathbf{K}_{t-i} \mathbf{H}_{t-i}) \right) (\mathbf{I}_m - \Phi) + \prod_{i=1}^{t-1} \Phi (\mathbf{I}_m - \mathbf{K}_{t-i} \mathbf{H}_{t-i}) \right] \boldsymbol{\lambda} \quad (13)$$

and

$$\text{bias}(\hat{\boldsymbol{\beta}}_{t|t}) = (\mathbf{I}_m - \mathbf{K}_t \mathbf{H}_t) \left\{ \left[ \mathbf{I}_m + \sum_{k=1}^{t-2} \prod_{i=1}^k \Phi (\mathbf{I}_m - \mathbf{K}_{t-i} \mathbf{H}_{t-i}) \right] (\mathbf{I}_m - \Phi) + \prod_{i=1}^{t-1} \Phi (\mathbf{I}_m - \mathbf{K}_{t-i} \mathbf{H}_{t-i}) \right\} \boldsymbol{\lambda}, \quad (14)$$

where  $\prod_{i=1}^n A_i = A_1 A_2 \dots A_n$ .

Thus, it is proved that induced forecast and filter bias are proportional to mean vector bias whose constant of proportionality is given by the expressions above. However, these expressions can be simplified to invariant models, i.e. when matrices  $\mathbf{H}_t$  do not depend on time.

## INVARIANT STATE SPACE MODELS

Let the model (1)-(2) represents an invariant state space model, i.e.  $\mathbf{H}_t = \mathbf{H}$  for all  $t$ , and the stationarity condition (3) holds. In this case, the Kalman filter converges fast to the steady-state Kalman filter.

Briefly, it means that the sequence  $\{\mathbf{P}_{t|t-1}\}$  converges to a steady matrix  $\bar{\mathbf{P}}_{t|t-1}$  which verifies the Riccati equation, and the sequence  $\{\mathbf{K}_t\}$  converges to a steady matrix  $\bar{\mathbf{K}}$  that verifies the equation  $\bar{\mathbf{K}} = \bar{\mathbf{P}}\mathbf{H}(\mathbf{H}'\bar{\mathbf{P}}\mathbf{H} + \boldsymbol{\Sigma}_e)^{-1}$ , ([5]).

For simplicity, it will be considered a univariate state space model ( $m = 1$ ). The limit of equation (13) when  $t$  goes to infinity and admitting that the steady Kalman gain  $\bar{K}$  is a good approximation of  $K_t$  for all  $t$ , is given by,

$$\begin{aligned} \lim_{t \rightarrow +\infty} \text{bias}(\hat{\boldsymbol{\beta}}_{t|t-1}) &= \lim_{t \rightarrow +\infty} \left[ (1 - \Phi) + \left( \sum_{k=1}^{t-2} \prod_{i=1}^k \Phi (1 - K_{t-i} H_{t-i}) \right) (1 - \Phi) + \prod_{i=1}^{t-1} \Phi (1 - K_{t-i} H_{t-i}) \right] \boldsymbol{\lambda} \\ &\approx \frac{(1 - \Phi)}{1 - \Phi(1 - KH)} \boldsymbol{\lambda} \end{aligned} \quad (15)$$

and

$$\begin{aligned} \lim_{t \rightarrow +\infty} \text{bias}(\hat{\beta}_{t|t}) &= \lim_{t \rightarrow +\infty} (1 - K_t H_t) \left\{ \left[ 1 + \sum_{k=1}^{t-2} \prod_{i=1}^k \Phi(1 - K_{t-i} H_{t-i}) \right] (1 - \Phi) + \prod_{i=1}^{t-1} \Phi(1 - K_{t-i} H_{t-i}) \right\} \lambda \quad (16) \\ &\approx \frac{(1 - \Phi)}{1 - \Phi(1 - KH)} (1 - KH) \lambda. \end{aligned}$$

Noting that in a steady state

$$KH = \frac{\bar{P}H^2}{\bar{P}H^2 + \sigma_e^2}, \quad (17)$$

that is  $0 < KH < 1$ , it can be concluded that bias of filter prediction are smaller than forecast bias. When  $H$  is large  $KH$  is approximately equal to 1, thus, in this case, filter and forecast bias are approximately zero and  $\lambda(1 - \Phi)$ , respectively. If  $H$  is small then  $KH$  is approximately zero and, in this case, both filter and forecast bias are equal to  $\lambda$ .

## ACKNOWLEDGMENTS

This work was partially supported by *FEDER* funds through *COMPETE*—Operational Programme Factors of Competitiveness (“Programa Operacional Factores de Competitividade”) and by Portuguese funds through the *Center for Research and Development in Mathematics and Applications* (University of Aveiro) and the Portuguese Foundation for Science and Technology (“FCT—Fundação para a Ciência e a Tecnologia”), within project PEst-C/MAT/UI4106/2011 with *COMPETE* number FCOMP-01-0124-FEDER-022690.

## REFERENCES

1. T. Alpuim, and I. Ribeiro, A state space model for run-off triangles, *Appl. Stoch. Model Bus.* **19**, 105–120 (2003).
2. C. F. Ansley, and R. Kohn, Prediction mean squared error for state space models with estimated parameters, *Biometrika* **73**, 467–473 (1986).
3. A. C. Harvey, *Forecasting, structural time series models and Kalman filter*, University Press, Cambridge, 1996.
4. J. D. Hamilton, A standard error for estimated state vector of a state-space model, *J. Econometrics* **33**, 387–397 (1986).
5. J. D. Hamilton, *Time series analysis*, Princeton University Press, Princeton, 1994.
6. D. Pfeiffermann, and R. Tiller, Bootstrap approximation to prediction MSE for state-space models with estimated parameters, *J. Time Ser. Anal.* **26**, 893–916 (2005).
7. R. H. Shumway, D. S. Stoffer, *Times Series and Its Applications*, Springer, New York, 2000.